Asymmetrical trajectories: The past and present of *–body/–one*

ALEXANDRA D'ARCY
University of Victoria

BILL HADDICAN
Queens College-CUNY

HAZEL RICHARDS
University of York

SALI A. TAGLIAMONTE
University of Toronto

ANN TAYLOR
University of York

ABSTRACT

The set of English [+human] pronominal quantifiers has been variable for at least 500 years, with the compound forms *–body* and *–one* competing since Middle and Early Modern English. This change has still to run its course (cf. Nevalainen & Raumolin-Brunberg, 2003:78). Using corpora of historical texts, we track the development of these variants alongside the demise of the earlier variant *–man*. Then, drawing on contemporary and regionally diverse corpora, we trace the continued development of *–body/–one* variation through the 20th century. The trajectories reveal paradigmatic leveling in the late 19th century and the rise of *–one* as the dominant form. However, grammatical, social, and lexical developments continue. Most striking is that after an initial phase of historical leveling, the different lexical

quantifiers—*any, every, some, no*—go their own ways in the collection of varieties examined here, demonstrating that the mechanisms shaping evolutionary pathways across the globe are not only systemic, but also retain local alternations.

INTRODUCTION

This paper focuses on variation between the affixes *–body* and *–one* in the pronominal quantifiers *somebody/someone*, *anybody/anyone*, *everybody/everyone,* and *nobody/no one*. Given that the *–body/–one* pairs behave similarly on standard diagnostics, most literature on the syntax and semantics of these pronouns considers them to be functionally equivalent ways of saying the same thing (see also Jespersen, 1914:444; Quirk, Greenbaum, Leech, & Svartvik, 1985:376–377; cf. Labov, 1972).[1] Both forms occur in analogous contexts within stretches of discourse of individual speakers in spontaneous speech, as in (1).

(1) a. I think he had *somebody* else as well as Betty anyway.
          But you have to live with *someone* to know them. (York)[2]
     b. No grief off *anybody*…It's your day, don't have to please *anyone* else. (York)
     c. Oh I think *everybody* knew someone that went, didn't they?…
          During the war *everyone* helped everyone else, didn't they? (York)
     d. There was *no one* to do it…*Nobody* likes my jumper. (York)

Despite the similar behavior of these alternants in contemporary English, their histories are quite distinct, deriving from a quantifier plus either the free morpheme *body* ('person') or cardinal *one*. How these forms emerged as variants during Middle English and the differences across lexical items have been traced through much of the Modern period.

   In this paper, we have a number of broad goals. The first is to track the historical development of *–body/–one* variation. Raumolin-Brunberg and her colleagues have provided a detailed picture of the development of the variants during Middle English (ME) and Early Modern English (EModE; e.g. Nevalainen & Raumolin-Brunberg, 2003; Raumolin-Brunberg, 1991, 1994; Raumolin-Brunberg & Kahlas-Tarkka, 1997). Their work draws primarily on the Helsinki Corpus of English Texts (Kytö, 1996; Rissanen, Kytö, & Palander-Collin, 1996) but is supported by a range of additional historical sources: for example, Corpus of Early English Correspondence ([CEEC], Nevalainen, Raumolin-Brunberg, Keränen, Nevala, Nurmi, & Palander-Collin, 1998), Century of Prose (Milić, 1990a, 1990b), Dictionary of Old English (Venezky & Healey, 1980), Shakespeare (Wells & Taylor, 1986). The current analysis draws on the Penn Historical Corpora to provide additional empirical evidence of the trajectories of change. These materials overlap with those used in previous research, but they crucially augment the window of analysis by 200 years. This expanded time frame reveals leveling of the lexical differences among the forms by the end of the 19th century.

The second goal is to tie the historical (British) evidence to the contemporary situation in the United Kingdom, drawing on a collection of synchronic corpora from across England and Scotland. The sum of these materials indicates that stylistic and social effects have been continuously significant predictors of variation.

The final goal is to perform a systematic comparative analysis across major varieties of English worldwide: British, American, Canadian, and New Zealand. An intriguing trend is visible across the datasets. There is ongoing change toward use of *–one*, but not for all quantifiers equally. Despite relatively stable social and stylistic effects, differentiation among the lexical quantifiers emerges across dialects.

THE PAST OF *–BODY/–ONE*: HISTORICAL CORPORA

Previous research

In Old English (OE), a broad array of forms semantically akin to the contemporary *–body/–one* quantifiers existed. There were two main syntactic types: (i) simple pronouns (e.g., *hwa* 'who', 'anyone', 'someone'; *hwelc* 'which'; *sum* 'some'; *ænig* 'any'; *nænig* 'none, no one'; *gehwelc* 'everyone'; *ælc* 'each'), and (ii) quantificational collocations with the indefinite pronoun *man* (e.g., *sum man* 'someone'; *ænig man* 'anyone'; *nan man* 'no one'). *An*, the ancestor of modern *one*, had a wide range of functions in OE as in Present Day English (PDE), as, for instance, a numeral (e.g., *one* as opposed to *two*), something approaching an indefinite article (PDE *a/an*), a particularizing/individualizing use (a certain person/thing), a partitive sense (one out of a group), as well as a number of minor types. It was not, however, used pronominally with reference to a particular or indefinite person, this function being supplied by *(ge)hwa* and *man* (Rissanen, 1967). Both the independent pronominal use of *one* and its use in compounds, as in (2), emerged in the 13th century; compounds with *–body*, exemplified in (3), followed about a century later. Both innovations competed with the traditional compound, as in (4). In conjunction with ongoing semantic shifts in the meaning of *man* and narrowing of its scope in other pronominal uses, this competition contributed to the obsolescence of the *man* compounds (Raumolin-Brunberg, 1994; Rissanen, 1997). In contemporary usage, the *man* form survives only in a limited number of set phrases and expressions, such as *no man's land*, *everyman's library*, and the well-known popular culture example in (5).

(2)   Philip answered him, Two hundred peny-worth of bread is not sufficient for them, that *eurey one* of them may take a litle. (1611, The New Testament, Authorised Version, VI)
(3)   A better *body* drank *neyuer* wine. (c1340 Cursor M. 3360 [Fairf.])
(4)   And sitting in some place, where *no man* shall prompe him, by him self, let him translate into Englishe his former lesson. (c.1568 Roger Ascham, Asch.e1-h:1v.22)
(5)   To boldly go where *no man* has gone before. (opening narration of *Star Trek*, 1966)

The innovative *–body/–one* compounds did not enter the language all at once. As outlined in Figure 1,[3] *–one* initially occurred with *every* and only gradually

FIGURE 1. Development of –body and –one, evidence from the Helsinki Corpus (based on Raumolin-Brunberg & Kahlas-Tarkka, 1997:75, Figure 4).

spread to the other quantifiers, while –body first appeared with no before generalizing. Generalization across a common set is a hallmark of grammaticalization (Heine, 2003:579–580; Hopper & Traugott, 2003:100–103), and it is relevant here because it predicts that the frequency of each form will differ across the individual lexical items, at least in the earlier stages of change. However, what has never (to our knowledge) been reported in the case of competing forms such as this is what happens as such a situation evolves. One expectation is that the variants might ultimately converge on the same pattern (e.g., via analogy). An alternative expectation is that the innovative forms might actually create the paradigm from their ongoing competition.

A picture of robust historical variation emerges from Raumolin-Brunberg (1994) and Raumolin-Brunberg and Kahlas-Tarkka (1997), who recorded a high degree of variability in the frequencies of –man, –body, and –one in both ME and EModE. However, use of –body increased rapidly after its introduction to the system. By the end of the EModE period, its frequency surpassed that of –one with all of the quantifiers except every (Raumolin-Brunberg, 1994; Raumolin-Brunberg & Kahlas-Tarkka, 1997:66, Figure 3), where –one had first emerged as a pronominal quantifier restrictor and remained well entrenched.

An intriguing question concerns why two variants should develop in all four sets over similar time periods (cf. Raumolin-Brunberg & Kahlas-Tarkka, 1997:76).[4] That is, despite having different temporal onsets and rates of development, the –body/–one compounds have arguably been analogous since their inception as pronominal quantifiers (even if not direct competitors until EModE). This suggests that linguistic predictors underlie their use. Additionally, consideration of social predictors reveals that stylistic and social conditioning were present early in the development of the innovative variants, if not from their inception.

In the Helsinki materials, the –body compounds were restricted to oral genres of writing (i.e., fiction, drama, private letters) until the middle of the 17th century (Raumolin-Brunberg, 1994:315–316; Raumolin-Brunberg & Kahlas-Tarkka,

1997:68). It was not until the end of the EModE period that they spread to literate genres. In contrast, the *–one* forms occurred across genres from the outset. Thus, *–body* appears to have had vernacular (or colloquial) origins, whereas *–one* was more ubiquitous. As we shall see, it was this advantage that may have laid the groundwork for its eventual success as the leading variant in contemporary usage.

New findings

The English historical corpora currently at our disposal enable us to probe further stylistic nuances in the use of the alternants. Register differences can be explored by comparing the more formal literary texts of the Penn-Helsinki Parsed Corpus of Early Modern English ([PPCEME], 1500–1710; Kroch, Santorini, & Diertani, 2004) with the less formal letters of the Parsed Corpus of Early English Correspondence ([PCEEC], 1410–1695; Taylor, Nurmi, Warner, Pintzuk, & Nevalainen, 2006). We can also consider levels of (in)formality within the letter corpus as measured by audience design (Bell, 1984), that is, the nature of the writer/addressee relationship.

Figure 2 displays the rate of *–body* over time according to register. To establish continuity with the contemporary data to follow, we calculate all historical rates using the binary contrast between innovative variants (i.e., *–body* and *–one*). The results illustrate that the variants are uniformly stratified across the Late ME and EModE periods: *–body* is consistently more frequent in the informal PCEEC than in the literary PPCEME. Thus, as *–body* spreads across the quantifiers, a stylistic distinction is maintained. Its occurrence in the literary materials is relatively infrequent until the end of the EModE period, at which time it increases in frequency in both genres.[5] These findings thus corroborate the results of earlier work: *–one* was the more prestigious variant, occurring more frequently in literate genres of writing, whereas *–body* functioned as the more vernacular form. However, it is critical to bear in mind the individual lexical trajectories of Figure 1. Increasing use of *–body* in Figure 2 is magnified by the late emergence of *–one* in the *no* and *some* sets, giving them inflated rates
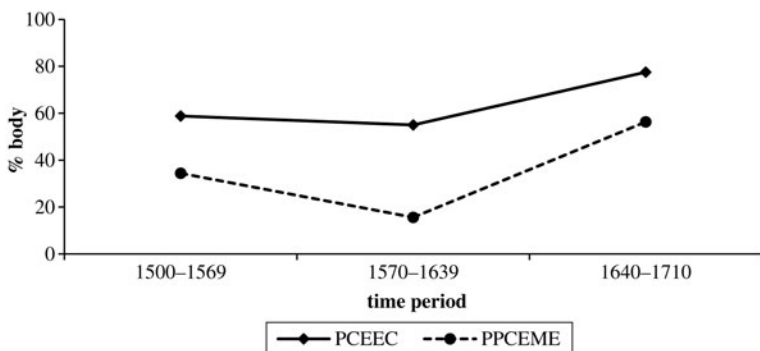

FIGURE 2. Distribution of *–body* over time by genre (letters, PCEEC; literary, PPCEME).

of *–body*. The other sets, in which *–one* developed early, remain majority contexts for *–one* in literary texts across all time periods.

The stylistic effect can be explored further within the PCEEC, where the social and/or emotional proximity of the writer to the intended addressee can be measured, offering insight into style shifting according to familiarity (see Bell, 1984:163). The letters were classified as belonging to one of three categories: "close" recipients, typically members of the nuclear family (mothers, fathers, siblings, and so on); "mid-distant" recipients, generally non-nuclear family, friends, and servants; and "distant" recipients, such as business associates and the like.[6] Figure 3 reports the results, revealing that as proximity increases, there is a concomitant increase in the use of *–body*, a pattern that is visible across the full set of quantifiers.[7] There is thus a negative correlation between formality and the *–body* variants, a correlation that is both strong and significant overall (r = −.999986, p = .003). This reinforces the idea that *–body* was the vernacular form, used in more casual situations. In contrast, *–one* correlates with formality and the expected traits of a standard variant.

The mounting historical evidence thus strongly suggests that the *–body*/*–one* alternants were stylistic variants. Given that style variation is generally held to derive from social meaning (see Bell, 1984:151), it is likely that social forces were also operative during this period. The CEEC materials, which represent a fairly colloquial mode of communication, provide support for sociosymbolic differentiation in the use of *–body*/*–one*. Men continued to use the traditional collocations with *–man* at robust rates into the EModE period. Women, however, led the shift away from these forms and they were marginal contenders from at least 1600 onward (Nevalainen & Raumolin-Brunberg, 2003:124). In their stead, women adopted the innovative forms, but not indiscriminately. Until the mid-17th century, women were well in advance of men with respect to *–body*, yet they lagged behind with respect to *–one*. In the second half of the century, as has been observed for other features in the CEEC (see Nevalainen & Raumolin-Brunberg, 2003:125–130), the gender profiles of the innovative variants
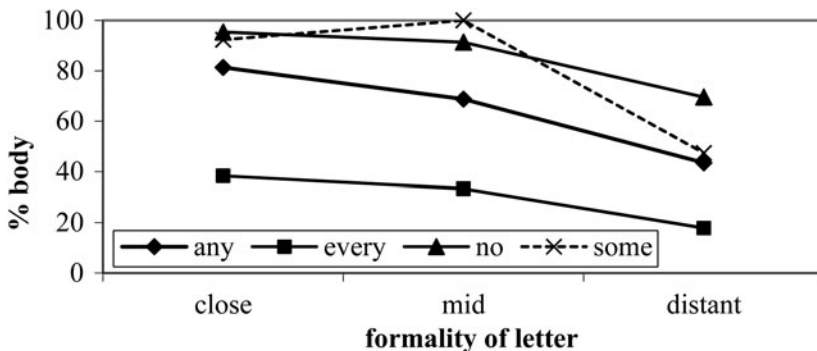

FIGURE 3. Distribution of *–body* in the PCEEC according to the proximity of writer and recipient (i.e., formality of letter), 1410–1695.

reversed. That is, according to the results reported in Nevalainen and Raumolin-Brunberg (2003:124, Figures 6.8a, 6.8b), by the year 1660, women had surpassed men in their use of *–one* while falling slightly behind them in the use of *–body*.[8]

The stylistic effect in the Helsinki and Penn materials and the shift in gender associations in the CEEC together suggest that by the EModE period social meanings constrained the choice of the innovative forms. By 1700, the *–one* forms bore the hallmarks of an emerging "standard" or prestige variant, associated with women and literate genres. In contrast, the *–body* forms were developing indications of a vernacular variant associated with men and oral genres. These systematic social constraints on *–body/–one* in the historical materials, therefore, reveal not only the entrenchment of the *–body/–one* quantifiers in the linguistic system, but also the secondary yet vital development of sociolinguistic meaning (cf. Eckert, 1999; Labov, 2001; Tagliamonte & D'Arcy, 2007).

The remaining issue concerns the ongoing development of the quantifiers in the Modern period. To this end, data from the Penn Parsed Corpus of Modern British English ([PPCMBE], 1700–1914; Kroch, Santorini, & Diertani, 2010) can be added to the perspective provided by the earlier corpora. The PPCEME and the PCEEC overlap in time (and, to some extent, content)[9] with the Helsinki Corpus and the CEEC, and so the findings reported for this period overlap with those already published by Raumolin-Brunberg (1994), Raumolin-Brunberg and Kahlas-Tarkka (1997), and Nevalainen and Raumolin-Brunberg (2003). The results for the Modern period, post-1700, are new.

Figure 4 tracks the frequency of *–body* from 1570 to the end of the 19th century. By 1700, the *–man* compounds were obsolescent and from that point forward, variation between *–body* and *–one* is binary. There is a high degree of variability between the different forms in the EModE data, reflecting their diachronic layering across time.[10] Consistent with its historical development (Figure 1), *–body* in this period is most frequent with the negative quantifier, *no*; as
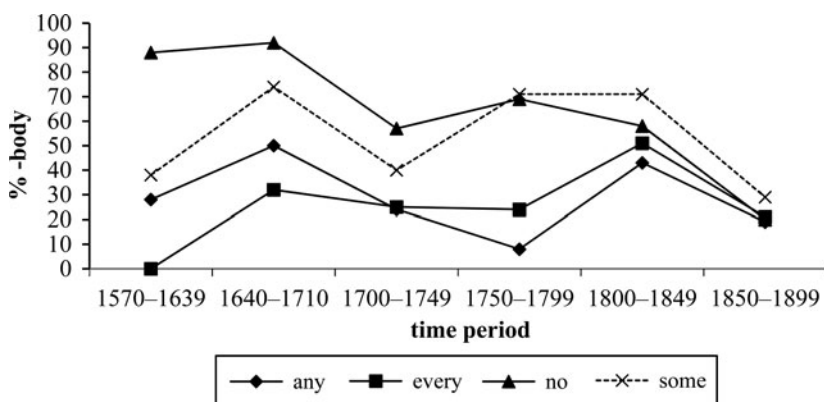

FIGURE 4. Distribution of *–body* by quantifier, 1570–1899 (PPCEME, PPCMBE).

reflected in Figure 4 by the exceptionally high rate of –*body* in the early periods, this was the final quantifier to host –*one*.

Raumolin-Brunberg and Kahlas-Tarkka (1997:77) framed their discussion of –*body*/–*one* alternation as "a development towards a tight and symmetrical repertoire of the indefinite pronouns." That is, an integrated category emerged in which each quantifier constitutes a target for –*body* and –*one*. The key insight added by the inclusion of data from 1850 to 1899 is that this development lead to "paradigmatic cohesion" (Lehmann, 1982:132–137): Earlier lexical differences (1570–1799) gradually yet consistently narrowed across time. First, *any* converged with *every*, followed by *no* and *some*. Around 1800, the paradigms began to pattern as a homogenous set, and by the end of the window examined here, 1899, any lexical effects in the distribution of –*body* had leveled. Thus, whereas the ME and EModE periods were characterized by robust lexical differentiation, Figure 4 reveals the eradication of lexical effects and the development of a coherent morphosyntactic category in the Modern period.

The trajectories in Figure 4 also reveal that despite the early overall prominence of –*body*, lexical leveling entailed a focusing toward –*one* in written English. Whereas there was an upswing in the use of –*body* for *every* and *any* in the period between 1800 and 1849, in the second half of the 19th century the rate of –*body* decreased markedly across the board.

The diachronic evidence thus reveals that once introduced, –*body* and –*one* ousted –*man*. The innovative forms behaved independently from one another until the 19th century, at which point paradigmatic leveling occurred, resulting in a cohesive set.[11] In historical written English, the –*one* compounds dominated. However, the two variants, –*body* and –*one*, had strong social correlates in historical written English, conditioned by style and author sex from the earliest time period, suggesting that –*body*/–*one* variation arose with intrinsic social meaning. Twentieth-century grammars record continuity of this meaning (e.g., the –*body* forms are considered the less elegant option; Jespersen, 1914:444; Quirk et al., 1985:376–378), suggesting that the sociolinguistic roots of –*body*/–*one* remain intact. However, it is yet unclear whether lexical leveling resulted in the entrenchment of a stable variable system or whether there is ongoing change. To address these questions, we first draw on a number of synchronic corpora from England and Scotland. We then consider New World Englishes.

THE PRESENT OF –*BODY*/–*ONE*: CONTEMPORARY BRITISH
DIALECTS

The contemporary British English data come from three regions of the United Kingdom: Scotland, Northern England, and the Midlands. The Scottish data were mined from the spoken components of the Scottish Corpus of Texts and Speech, henceforth the Scots Corpus.[12] The Northern English data come from the York Corpus (Tagliamonte, 1996–1998) and the Newcastle recordings in the

TABLE 1. *Contemporary British dialect data*

Location	Collection dates	Birth years, younger speakers	Birth years, older speakers	Tokens, n
Scotland	2002–2006	1970–1989 ($n = 40$)	1920–1949 ($n = 15$)	355
Newcastle	1994	1967–1979 ($n = 17$)	1927–1949 ($n = 14$)	345
York	1994–1996	1965–1980 ($n = 19$)	1925–1950 ($n = 37$)	718
Derby	1995	1968–1981 ($n = 14$)	1928–1950 ($n = 15$)	221

Phonological Variation and Change materials (Milroy, Milroy, & Docherty, 1997). This latter collection also provided the Midland data, as it includes recordings from Derby. These materials provided 1639 tokens of *–body/–one* pronouns; Table 1 lists the details.[13]

We examine social and linguistic effects on *–body/–one* by fitting a generalized linear mixed effects regression model, with *–body/–one* as the dependent variable and a random intercept for speaker.[14] The fixed social predictors are *age group*, *sex*, *socioeconomic status* (SES), and *corpus* (region). The first three predictors each have two levels: older vs. younger; men vs. women; and professional vs. nonprofessional.[15] The fourth predictor has four levels, one for each corpus.

The fixed linguistic predictors are *quantifier*, with levels *any*, *every*, *some*, and *no*, and *postnominal modifier*, with levels present vs. absent. A key distinction between pronouns (e.g., *somebody*) and quantified determiner phrases (e.g., *some boy*) is the ability of the former to take a postnominal adjective (*somebody nice* vs. *some boy nice). The modifier predictor thus tests for the possible effect of a postnominal adjective, as demonstrated in (6), though modification largely reflects a single token type, *else*, as in (6c) and (6d).[16]

(6)   a.   Gatenby's *someone different*, obviously. (York)
      b.   Nearest (inc) we have to *anybody exotic* at the moment is Laura, from Romania. (York)
      c.   *Everyone else* ran away and left me to talk to him. (Newcastle)
      d.   *Nobody else* would volunteer. (Derby)

Table 2 lists the results for the predictors selected in the model.[17] Five fixed effects are reported: SES with the reference level "nonprofessional," age group with the reference level "older," modifier with reference level "none," and corpus with the reference level "York." In addition, the model includes interactions between age group and modifier, and age group and corpus.

The intercept provides a baseline from which the model predictions are built. In this case, it is highly positive (2.294 in log odds units), and a significant reading ($p \leq .0001$) shows the strong tendency in the data toward *–body* overall. The model also illustrates the relative influence of the social and linguistic predictors. Whereas speaker age overall is not significant, notice the impact of the interaction terms with speaker age and modifier ($p \leq .0001$) and speaker age

TABLE 2. *Coefficients (logits), standard errors, z values, and p values for fixed effects in the combined model, with 1 ("application value") = –body*
n observations = 1639; n speakers = 171
Akaike information criterion: 1421; Bayesian information criterion: 1502
logLik: –695.5
deviance: 1391

Random effects					
Group	Variance	SD			
Speaker (Intercept)	4.7967	2.1901			
Fixed effects					
Group	Estimate	SE	z value	Pr(>|z|)	n
(Intercept)	2.294	.481	4.772	$1.82e^{-06a}$	1108
SESProfessional	−1.278	.457	−2.798	$.0051^{b}$	1022
AgeYounger	−.558	.696	−.802	$.4223^{e}$	880
ModifierYes	−1.044	.361	−2.893	$.0038^{b}$	213
CorpusScots	1.793	1.000	1.793	$.0730^{d}$	355
CorpusDerby	1.400	.902	1.552	$.1207^{e}$	221
CorpusNewcastle	3.274	1.149	2.850	$.0044^{b}$	345
QuantifierAny	.183	.236	.776	$.4377^{e}$	254
QuantifierEvery	.360	.188	1.918	$.0551^{d}$	527
QuantifierNo	.743	.248	2.992	$.0028^{b}$	233
AgeYounger:ModifierYes	1.596	.459	3.476	$.0005^{a}$	125
AgeYounger:CorpusScots	−3.046	1.220	−2.498	$.0125^{c}$	227
AgeYounger:CorpusDerby	−2.771	1.254	−2.210	$.0271^{c}$	136
AgeYounger:CorpusNewcastle	−5.411	1.405	−3.852	$.0001^{a}$	226

Notes: Significances are as follows: [a]0, [b]0.001, [c]0.01, [d]0.05, [e]0.1.

and corpus, which underscores an extreme difference between York and the other corpora, foreshadowed in the corpus effect overall. This suggests linguistic change in progress involving differential patterning geographically across British dialects, as well as an internal linguistic effect involving the quantifiers and their modifiers.

These latter findings lead us to probe the interaction of age and regional effects. We illustrate the relationship between these predictors in the partial effects plot in Figure 5 (in this plot and in other partial effects plots to follow, log odds have been converted to probabilities).

Figure 5 shows a dramatic picture of consistency but also of systematic differentiation across the communities. Younger speakers tend toward *–one* in all of the corpora (the dashed line), but in the Scots Corpus, Derby, and Newcastle this difference is increasingly marked. It appears that younger people are converging on the *–one* forms, but they are doing so at a different pace by community. Why are the younger people in all these communities using more *–one* than their elders are? There are two possible interpretations. One is that this variable is age-graded: younger speakers use more *–one*; however, this is a passing phase. Another is that these data reflect change in progress: *–body* is giving way to *–one* in the community grammars, consistent with the trends in the historical data. One argument in favor of this latter view is theory-internal. Variation between *–body* and *–one* is neither highly salient nor imbued with
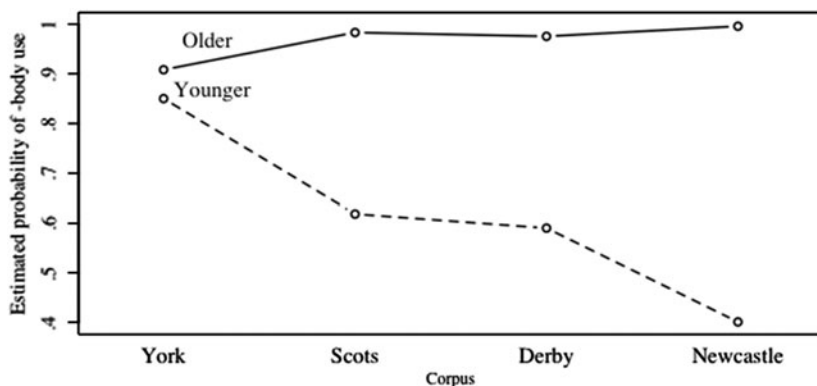
FIGURE 5. Estimated probabilities for the interaction between corpus and age group.
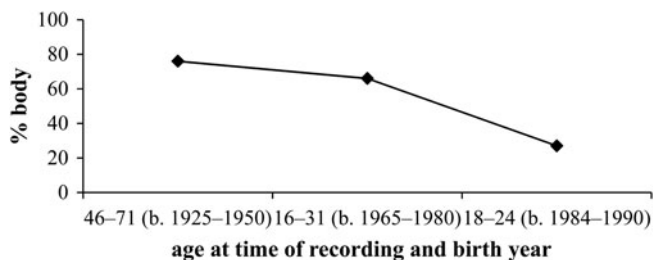

FIGURE 6. Distribution of –*body* in York among speakers born 1925–1990.

obvious social meaning, factors that are said to be typical of age-grading (e.g., Chambers, 1995). It does not appear, for example, to be subject to metalinguistic commentary by speakers (see Milroy & Gordon, 2003:36–37; also, Cheshire, 2005). However, when we probe the results obtained from the York data, in Figure 6, an empirical argument comes to the fore.

The left two speaker cohorts in Figure 6 illustrate apparent time results from Tagliamonte's York Corpus, collected from 1994 to 1996. The cohort on the far right comes from a similarly constructed sample, collected in 2008.[18] A decline in the use of –*body* is visible across the whole of this figure. The real-time decrease, from 66% overall in 1996 among speakers born between 1965 and 1980 to 27% overall in 2008 among speakers born between 1984 and 1990, both reflects and significantly advances the apparent time trajectory (χ^2 $p < .0001$), a trajectory that typically underestimates the rate of change. The more recent findings also bring the York distribution into alignment with that of the other corpora, suggesting that York is simply a little more conservative, a finding that echoes other research on this variety (e.g., Tagliamonte, 2006; Tagliamonte & Roeder, 2009). The overarching trend emerging from the analysis then is that younger speakers are moving away from –*body* and toward –*one* in informal

contemporary British English dialects, just as is documented in the historical record, albeit at different rates.

At the same time, the results in Table 2 and Figure 5 indicate that overall, –*body* (not –*one*) remains the more frequent form. However, this synchronic evidence comes from informal spoken discourse. The diachronic evidence presented earlier, where –*one* prevails, comes from written materials. As such, the greater frequency of –*body* in the contemporary materials may straightforwardly derive from the longstanding stylistic effect. If such is the case, then we can extrapolate that paradigmatic leveling was not specific to a particular genre (writing vs. speech), but that it applied across the board as a generalized aspect of the English [+human] pronominal quantifiers.

The question of style can be explored further if we marshal the evidence from an additional source of evidence, the British National Corpus ([BNC]; BNC Consortium, 2007), which contains both spoken and written materials (collection years: 1991–1994). To assess written norms, the newspaper section was targeted. The BNC yielded 23,021 tokens across genres (*n* spoken = 7434; *n* written = 15,587). The overall distribution of –*body* in each of the lexical sets is displayed in Figure 7, where—with the possible exception of *no*—there is no lexical effect. Crucially though, the BNC data reveal clear and systematic stratification of –*body*/–*one* in the expected direction: –*body* is consistently more frequent in speech, and –*one* is consistently more frequent in texts. Chi-square tests confirm that, for each quantifier, this difference is significant (all *p* values <.0001). We interpret this as confirmation that the stylistic distinction observed in the historical data is maintained in current usage (cf. Jespersen, 1914:444; Quirk et al., 1985:376–378) and that the high frequency of –*body* in the contemporary British dialect data derives from its spoken nature—a contemporary reflex of a longstanding diachronic pattern.

Returning then to the results in Table 2, note the significant interaction between age group and postnominal modification (AgeYounger:ModiferYes $p \leq .0005$). We illustrate this interaction in Figure 8. A following adjective (generally *else*) favors –*body* slightly for younger speakers, but disfavors –*body* among older speakers. We will return to this emergent effect in our discussion of New World Englishes.
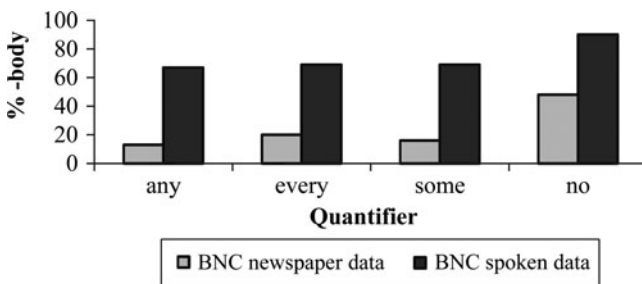

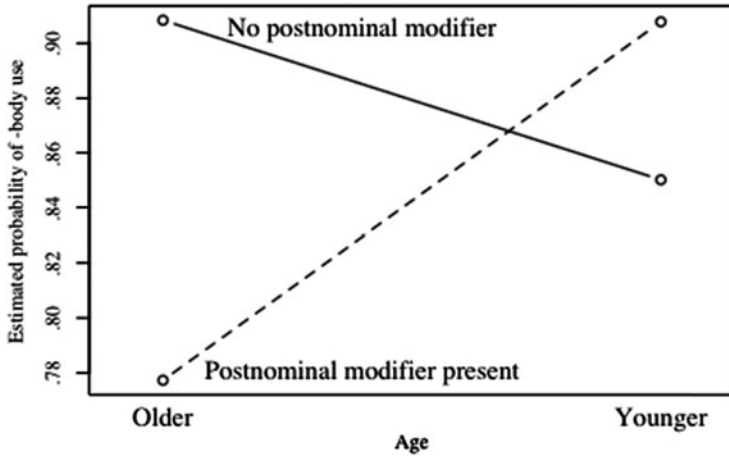FIGURE 7. Distribution of –*body* by genre in the BNC.

FIGURE 8. Estimated probabilities for –*body* by postnominal modifier and age group.
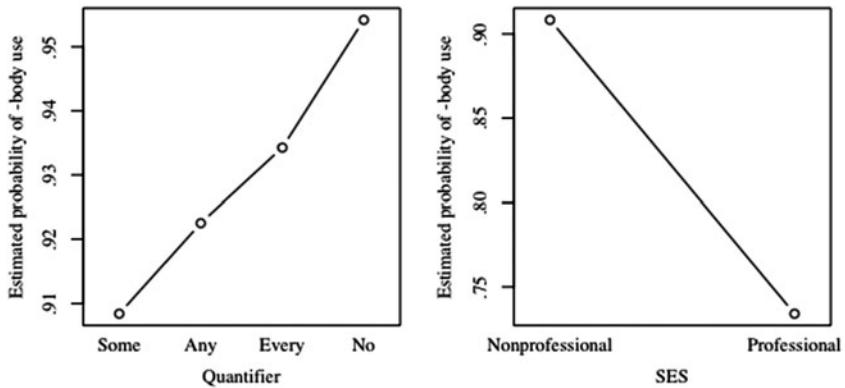

FIGURE 9. Estimated probabilities for quantifier (left) and speaker SES (right).

From the perspective of the historical data, the effect of quantifier in Table 2 is also noteworthy. The left panel of Figure 9 plots partial effects for the quantifier predictor, where it is apparent that *no* favors –*body,* but *some* disfavors it. Although the *no*/*some* contrast is the only significant one in the model reported in Table 2, this result suggests that despite the view of leveling from the second half of the 19th century, some of the historical lexical effects remain: *no* was the final quantifier to which –*one* generalized. This would seem to suggest that—as a consequence of historical persistence—*no* lags behind in the ongoing change, a result also supported by the BNC results of Figure 7.

Finally, the right panel of Figure 9 illustrates the effect of speaker SES. Nonprofessional speakers favor –*body,* and professionals favor –*one*. This result

aligns with the stylistic pattern in which –*one* is more frequent in both contemporary written texts and in historical formal written genres. In contrast with the (written) diachronic picture, however, there is no evidence of a speaker sex effect in the contemporary (spoken) corpora. Women show slightly higher rates of –*body* than men do (69% vs. 65%), but this difference is not significant overall.

To summarize, our analysis of the UK corpora has revealed three main results. First, there is evidence for change in progress toward –*one*. A real-time comparison with a similarly constructed sample in a single locale, York, confirms a real-time advance of –*one* (obviating an interpretation of age-grading). Second, the data are generally in keeping with the historical evidence in suggesting a leveling of the lexical effect on variation. In the spoken corpora, *no* favors –*body* relative to *some*, but no other contrasts between the quantifiers emerge. Third, there is an overarching commonality across dialects and age groups in the linguistic and social effects on –*body*/–*one* variation. That is, other than the developmental effect of postnominal modification (to which we return shortly), the significant interactions with age and variety are consistent with an interpretation of differential rates of change rather than dialect differentiation. We consider these observations further in the following section.

THE PRESENT OF –*BODY*/–*ONE*: THE NEW WORLD
PERSPECTIVE

Morphosyntax is held to be fairly homogenous across standard varieties of English worldwide (Görlach, 1991:25). In contrast, the sociolinguistic embedding of variable forms can be quite distinct from one locale to the next (Buchstaller & D'Arcy, 2009:317–320; Tagliamonte & Hudson, 1999:167). In the case of –*body*/–*one*, there are claims of regional differentiation in the literature, but these are sometimes contradictory (e.g., Bolinger, 1976, states that –*one* is more frequent in American English, whereas Biber, Johansson, Leech, Conrad, & Finegan, 1999, state that –*body* is more frequent in that variety). As demonstrated in the previous section, within the United Kingdom there is considerable regional parallelism. To date, however, systematic cross-variety comparisons of –*body*/–*one* variation have not been carried out. Given the findings presented in the previous section, it is worth asking which (socio) linguistic properties hold constant across other varieties of English. Given our assertion, based on the British dialect data, that longitudinal change is ongoing, the perspective of New World varieties should offer important insights.

We therefore add three additional synchronic datasets to the analysis, extending the geographic envelope to North America, Canada and the United States, and to the southern hemisphere, New Zealand.[19] Canadian English (CanE) is represented by the Toronto English Archive (Tagliamonte, 2003–2006). American English (AmE) is represented by the Buckeye Corpus (Pitt, Dilley, Johnson, Kiesling, Raymond, Hume, & Fosler-Lussier, 2007). New Zealand

TABLE 3. *Contemporary New World data*

Variety	Collection dates	Speakers, n	Ages and/or birth years	Tokens, n
NZE	1994–2007	165	1918–1988	1294
CanE	2003–2006	45	1912–1986	823
AmE	1999–2000	19	≤ 30 and ≥ 40 years old	699
BrE	1994–2006	171	1920–1989	1639

English (NZE) is represented by the Canterbury Corpus, the synchronic component of the Origins of New Zealand English Archive ([ONZE]; Gordon, Campbell, Hay, Maclagan, Sudbury, & Trudgill, 2004; Gordon, Maclagan, & Hay, 2007). All three are large urban samples, and each is representative of, but not tantamount to, broader national norms. However, where the Toronto and Buckeye data were each collected within a circumscribed timeframe (ca. 2004 and ca. 2000 respectively) and within a circumscribed locale (Toronto and Columbus, respectively), the Canterbury Corpus is a monitor corpus with no regional restriction (though all recordings are made in Christchurch and its local surrounds); the data were collected over a period of 14 years. Finally, although it is possible to test for socioeconomic effects in the Toronto and Canterbury datasets, the Buckeye data are restricted to middle-class speakers (Kiesling, Dilley, & Raymond, 2006:3).

Following the procedure for the UK samples (Table 1), speakers in the New Zealand and Canadian samples were assigned to age groups by birth year. Individuals born before 1951 were coded as "older" and those born after 1964 were coded as "younger." The Buckeye Corpus classifies speakers as older (>40 years) and younger (<30 years), meaning that the older speakers were born in or before 1959 and younger speakers in or after 1969. This age grouping aligns only partly with that used for the UK (British English [BrE]), Canadian, and New Zealand corpora. We summarize these samples in Table 3.

To compare linguistic and social effects on –*body*/–*one* in these New World datasets with those from the United Kingdom discussed previously, we fit a generalized linear mixed effects regression model with all the corpora. We recoded the corpus predictor, grouping the four BrE datasets together. The reconfigured predictor has four levels: NZE, CanE, AmE, and BrE. Because the Buckeye Corpus is not socioeconomically stratified, we do not include the SES predictor in this model. Instead, we model these effects in a separate run which excludes the AmE data. The predictors examined are otherwise the same as those described for the UK corpora, and the variable selection procedure was also the same.

Table 4 lists the results.[20] There are three fixed main effects: quantifier with the reference level *every*, age group with the reference level "younger," and corpus with the reference level "New Zealand." In addition, it includes interactions between quantifier and age group, and quantifier and corpus. Age aside, there is

TABLE 4. *Coefficients (logits), standard errors,* z *values and* p *values for fixed effects in the combined model, with 1 ("application value") = –body*
n observations = 4455; n speakers = 400

Fixed effects	Estimate	SE	z value	Pr(>|z|)	n
(Intercept)	− 2.289	.243	− 9.406	$<2e^{-16a}$	2602
QuantifierSome	.007	.243	.030	$.9762^e$	1594
QuantifierAny	.953	.311	3.069	$.0021^b$	641
QuantifierNo	.531	.316	1.681	$.0928^d$	612
AgeOlder	2.948	.287	.274	$<2e^{-16a}$	1872
CorpusAmE	2.804	.504	5.561	$2.69e^{-08a}$	699
CorpusCanE	2.818	.454	6.200	$5.64e^{-10a}$	823
CorpusBrE	2.366	.325	7.270	$3.61e^{-13a}$	1639
QuantSome:AgeOlder	− .645	.245	− 2.634	$.0084^b$	702
QuantAny:AgeOlder	− .636	.307	− 2.070	$.0384^c$	309
QuantNo:AgeOlder	− .809	.310	− 2.610	$.0090^b$	281
QuantSome:CorpusAmE	− .211	.372	− .568	$.5701^e$	269
QuantAny:CorpusAmE	− .632	.482	− 1.312	$.1895^e$	106
QuantNo:CorpusAmE	− .217	.522	− .415	$.6783^e$	82
QuantSome:CorpusCanE	− 1.075	.326	− 3.299	$.0010^a$	253
QuantAny:CorpusCanE	− 1.013	.415	− 2.441	$.0147^c$	107
QuantNo:CorpusCanE	− .342	.405	− .843	$.3991^e$	122
QuantSome:CorpusBrE	− .170	.293	− .582	$.5607^e$	625
QuantAny:CorpusBrE	− .893	.376	− 2.372	$.0177^c$	254
QuantNo:CorpusBrE	.121	.382	.315	$.7527^e$	233

Notes: Significances are as follows: [a]0, [b]0.001, [c]0.01, [d]0.05, [e]0.1.

great constancy of the social predictors across dialects. There are no interactions between corpus and age group, nor with speaker sex.

In contrast to the BrE dialect data (Table 2), the intercept is negative (–2.289 in log odds units), reflecting the strong tendency in the data toward *–one* overall. The model also exposes the relative influence of the social and linguistic predictors. The most significant of these by far is speaker age, whose *p* values are extremely low. The corpus (i.e., variety of English) is also highly significant. Each is differentiated from NZE with *p* values of <.0001, exposing the split between North America and the southern hemisphere. Of particular note are the widely varying varietal patterns visible in the interaction of quantifier and corpus. Note in particular quantifier *some* in the Toronto data (QuantSome:CorpusCanE), which is set apart from all the others, especially *every*, where the log odds units are –1.075 and the *p* value is low (*p* ≤ .0001). In contrast, in BrE *any* is set apart from *every* but not the others, and in AmE there is no statistically significant difference across the quantifiers. Finally, the analysis reveals no significant main effects or interactions for following modifier.

The main results in Table 4 are illustrated in the partial effects plots in Figures 10 and 11, which show the interactions between quantifier and age group, and quantifier and corpus, respectively.

Figure 10 shows an age effect familiar from the UK corpus results discussed previously. Younger speakers tend strongly toward *–one,* whereas older speakers
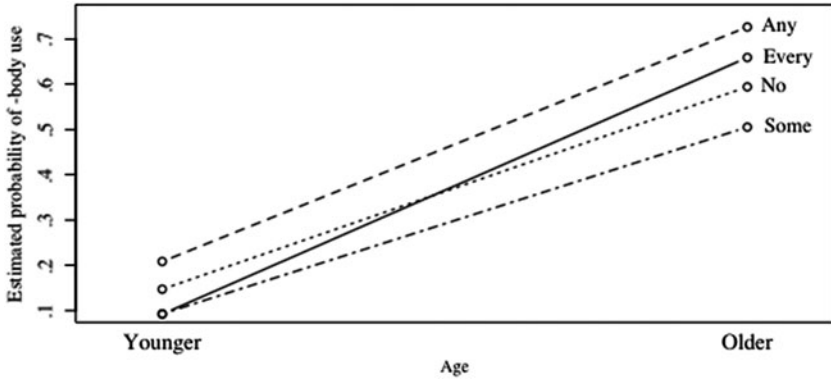
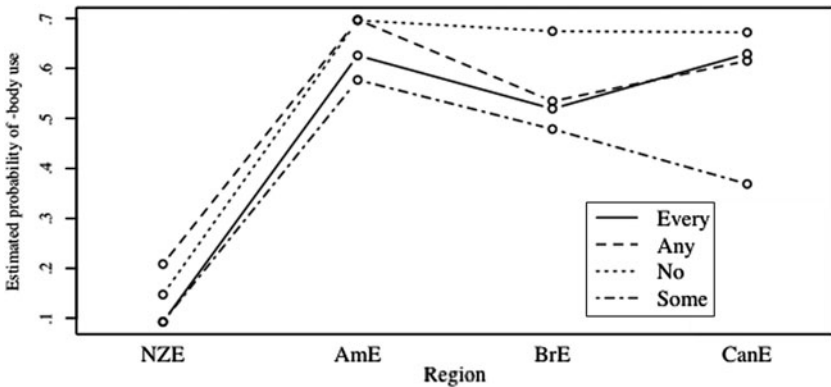FIGURE 10. Estimated probabilities for –*body* by quantifier and age group.

FIGURE 11. Estimated probabilities for –*body* by quantifier and variety.

retain greater use of –*body*. The analysis reported in Table 4 reveals no significant interactions between age group and corpus. These results suggest that the age effect is constant across the four varieties represented in the model and provide apparent time support for ongoing change toward –*one* across different varieties of English. This latter finding reflects the gradual shift evident in the historical written data (Figure 4), where use of –*body* decreased in real time as the system focused on –*one*. The results also show an interaction within the quantifiers. Among younger speakers, *every* patterns with *some*, disfavoring –*body*, but among older speakers, the coefficient for *every* is greater than are those for *no* and *some*. These findings reveal that as this change progresses, the quantifiers are converging, yet some differentiation remains among them.

Figure 11 plots partial effects for quantifier and corpus, and the results offer a dramatic portrait of dialect differentiation. The interactions reveal local deviations in the effect of quantifier on the choice of –*body* vs. –*one*.[21] They

also reveal that regardless of quantifier, New Zealand speakers show far greater use of *–one* (the probabilities for *–body* are exceptionally low). In contrast, the North American and UK English samples show significantly greater use of *–body*.[22] One possible interpretation of this difference is that NZE is more innovative than are the other varieties. Alternatively, New Zealand speech has strong roots in the varieties of London and southeast England (Gordon et al., 2004:46–47). If, as suggested by Bolinger (1976), *–one* is associated with southern UK dialects, then the NZE penchant for *–one* may be a settlement legacy. Mitigating strongly against this interpretation, however, is the perspective provided by the full ONZE Corpus (Gordon et al., 2004). Use of *–one* is infrequent among the first generation of native speakers, born in 1850 and the decades immediately following (Mobile Unit: 35.2% overall, $n = 216$). Use of *–body* remained relatively frequent overall until the 1930s and 1940s, when it began a period of rapid decline.

As we have already discussed, because the Buckeye Corpus includes no SES stratification, the model in Table 4 does not include this predictor. To compare the effect of SES in the other samples, we fit a separate model excluding the AmE data. The selected predictors match those of Table 4, with the addition of speaker sex as a main effect. There were no significant interactions with speaker SES. Figure 12 plots the estimated probabilities for speaker SES and shows that CanE and NZE pattern in lock-step with BrE. Nonprofessionals consistently favor *–body,* whereas professionals consistently favor *–one*. This social effect thus operates tenaciously across these major varieties of English, entrenched regardless of how far the overall shift toward *–one* has advanced.

To review, the analyses for the contemporary corpora, UK and New World, support two important findings. First, across the seven corpora, there is a strong age effect. Older speakers tend toward *–body,* whereas younger speakers tend toward *–one*. The convergence of apparent and real-time evidence suggests that this represents an overarching change in progress toward the *–one* variants.
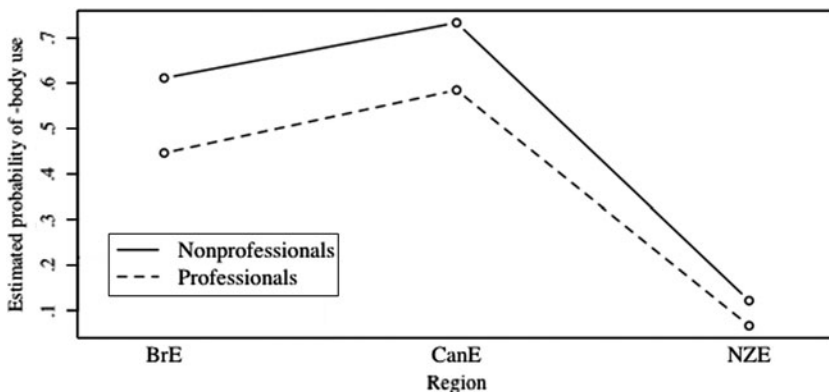

FIGURE 12. Estimated probabilities for *–body* by corpus and SES.

Second, the models reveal a contrast between linguistic and social predictors across dialects. In each locale, there is a different hierarchy of forms across quantifiers. In NZE, the quantifiers behave similarly in favoring –*one*; in AmE, the quantifiers behave similarly in continuing to favor –*body* (despite the trajectory toward –*one*); in CanE, *some* is set apart, favoring –*one*; in BrE, it is *no* that is set apart, favoring –*body*. At the same time, the social effects are parallel in each dataset. Except for the interaction between age and corpus in Table 2, which we interpret as reflecting a slower rate of change in York (Figure 2), corpus does not interact with any of the social predictors included in our models (age group, sex, SES). Next, we will discuss some broader implications of these results.

DISCUSSION: NEITHER ECONOMY NOR SYMMETRY

–*Body*/–*one* presents a good example of a variable that has persevered for many centuries—at least 500 years for certain quantifiers. As a consequence, this variation offers a unique perspective on linguistic change. The variants are syntactically and semantically equivalent and are apparently to this day unremarked outside of grammar books. Historically, at least two of the quantifiers (*any* and *every*) were used with –*body* more frequently in letters than in formal literary texts, and where these are employed in letter writing, they are more common in letters to close recipients (such as immediate family) than they are in letters to more distant correspondents (such as business associates). Synchronically, this stylistic effect is reflected in the lower rate of –*body* in written versus spoken data in the BNC (Figure 7) and is mirrored by the higher use of –*body* among nonprofessionals. The stability of the social distinction between –*body* and –*one* may, in part, have prevented the quantifiers from resolving upon one variant or the other. Thus, it appears that ingrained stylistic nuances have maintained long-term variation.

At the same time, we have documented an incremental, longitudinal change. At the onset of the ME period, the preceding quantifier strongly constrained selection of –*body* vs. –*one* (1570–1639). Over 300 years, the quantifiers gradually converged (1850–1899, Figure 4), suggesting "paradigmatic cohesion." From that point onward, there is an ongoing shift toward –*one*, as reflected in the consistent differences between older and younger speakers in the contemporary data (Figure 10) and in the real-time confirmation from York (Figure 6). Given these congruent findings, we could logically hypothesize that the change toward –*one* may eventually go to completion, opting for economy of form. However, in reality, the effect of the preceding quantifier varies considerably across varieties of English worldwide (Figure 11). Indeed, virtually every possible outcome obtains. In some varieties, there is continued cohesion, yet each favors a different form (–*one* in NZE and –*body* in AmE). In the other varieties, one quantifier or the other favors a form in contrast to the rest of the paradigm (*someone* in CanE; *nobody* in BrE). Thus, it appears that linguistic evolution is only systemic to a point. Local conditions offer opportunities for divergence within the same (variable) system.

CONCLUSIONS

The four [+human] indefinite pronominal quantifiers (*any*–, *every*–, *no*–, and *every*–) of contemporary English originated in the demise of an earlier form (–*man*), which was first encroached upon by –*one*, followed by –*body*. This led to a long period of variation founded in stylistic conditioning. In time, –*man* fell away; –*body* and –*one* competed; and by the 19th century, –*one* prevailed across all the quantifiers in writing, leading to a linguistic explanation of paradigmatic leveling (Figure 4). Change, however, did not stop there.

Variation in the forms used with the quantifiers endures in spoken English. In all of the synchronic datasets examined here (United Kingdom: Scotland, Newcastle, York, Derby; New Zealand: Canterbury; North America: Toronto, Buckeye), the quantifiers exhibit a continued shift toward –*one*. Given that –*one* is the more prestigious of the two variants (historically and contemporaneously), standardization seems an uncontroversial hypothesis, even if the intention of speakers is not necessarily standardization in and of itself. The consistency in the nature (recession of –*body*) and the timing of when the shift appears to have accelerated (ca. 1930s) across geographically and socially diverse locales, in conjunction with the mundane, quotidian nature of the variation itself, supports this interpretation.

Yet even after hundreds of years, the quantifiers have not shifted to categorical use of –*one* and thus a wholesale move toward economy and symmetry does not obtain. Instead, the varieties of English represented in our sample go their own way, not only in terms of form but also in terms of the distribution of the form across the quantifiers. Thus, although "unplanned purposefulness" (Keller, 1989:113)—the idea that language proceeds of its own accord—may be the mark of evolution of linguistic systems in an idealized universe, the results of our analysis expose a fine-grained array of differentiation in the real world. We suggest that the nuanced cross-variety portrait in English –*body*/–*one* variation provides a potent reminder that local linguistic ecologies play a critical role in shaping the practical eventualities of language use in the speech community.

NOTES

1. There are a number of diagnostics that establish the –*body*/–*one* forms as syntactically equivalent (as opposed to quantifier + noun phrase [NP] sequences). First, they both take a postnominal modifier, whereas quantifier + NP sequences cannot (Kishimoto, 2000; Larson & Marusic, 2004): I spoke to *somebody/one interesting* at the party vs. *I spoke to *some boy interesting* at the party. Second, whereas quantifier + NP sequences can be plural, the –*body*/–*one* forms are morphologically singular: I met *some boys* at the party vs. *I met *somebodies/ones* at the party. Third, unlike quantifier + NP sequences, the –*body*/–*one* forms can be the subject of imperatives (Zanuttini, 2008): *Nobody/one move*! vs. **No boy move*! Fourth, when affixed to *every*, the pronouns can support collective readings, whereas *every* + NP cannot: *Everybody/one* lifted the table together vs. **Every girl* lifted the table together. Finally, in contrast to *every* + NP, for many speakers *every* + *body/one* allow dependent plural readings (at least marginally): *Everybody's/one's noses* are made of cartilage vs. **Every girl's noses* are made of cartilage. On syntactic doublets in general, see Aronoff (1976), Embick (2008), Kroch (1994), and Taylor (1994).

2. Parenthetical information following examples contains citation information; examples from contemporary speech corpora indicate the source region/corpus.

3. This figure does not show any results for *no one*. This is because there are no examples of this form in the Helsinki Corpus; however, Shakespeare's texts do provide evidence for its presence in the language around 1600 (Raumolin-Brunberg & Kahlas-Tarkka, 1997:74).

4. This question becomes even more intriguing when considered in cross-linguistic perspective. As far as we can tell, it seems that of the West Germanic languages, English may be the only one to make compounds with *–body*. For example, High German, Low German, Dutch, Scots Gaelic, and Frisian use both *–one* and *–man*, where *man* is sometimes 'person' (Frisian *elkenien*, lit. 'every + one' and *immen*, lit. 'some + people/men'). Afrikaans uses just *–man* (*niemand*, lit. 'no + man'), whereas Yiddish uses *–one* (*keiner nit*, lit. 'one + no'). Nonetheless, competition within the paradigm of indefinite quantifiers appears to be somewhat rare and restricted. Dutch, Afrikaans, and Yiddish do not appear to exhibit any form/function asymmetry for these quantifiers. In West Frisian, there is variation between *nimmen* and *gjinien* for 'nobody', and between *immen* and *ien* for both 'anybody'/'somebody'. In High German, there is variation between *keiner* and *niemand* for 'nobody', and between *irgendeiner* and *jemand* for 'anybody'. English thus seems to be the only language in the West Germanic family to exhibit parallel competition across the whole of the paradigm, and it has done so with historically novel forms. Why this should be the case is both perplexing and worthy of further scrutiny, but this avenue is beyond the scope the current analysis. We thank an anonymous reviewer for pointing us in this direction.

5. Unlike the Helsinki Corpus, *–body* is not avoided entirely in the literary materials, but the PPCEME may well contain materials that would qualify as oral in Raumolin-Brunberg's system of categorization.

6. One reviewer rightly queried whether hierarchical relations existed within some of these categories, particularly in the case of servants. However, as there are just 10 tokens in the servant subgroup (2 as writer, 8 as recipient), the available data are insufficient for probing this possibility.

7. The exceptional value of *some* in the "mid" familiarity category derives from low token numbers ($n = 3$).

8. It is interesting to speculate why this reversal took place at this time, bringing the sexes into alignment with more general patterns of language change (e.g., Labov, 1990). Although we could hypothesize as to causation (e.g., the changing status of women and their concomitant literacy many have led women to a greater use of formal styles of writing), we are not in a position to fully explain this result at this time.

9. The PPCEME contains significantly more text than does the Helsinki Corpus of the same period, and the PCEEC contains slightly less than the CEEC.

10. In this figure, *some* is represented with a dotted line because of the comparatively small amount of data available when separated by time period. Consequently, the number of tokens in some periods is so small (e.g., $n = 5$) that the patterns cannot be taken as meaningful.

11. We are aware, however, of at least two exceptions, both involving *–one*. Some Scottish English dialects have pronominal forms with the quantifier *all* (typically reduced to [a]), including *athing* 'all thing' and *abody* 'all body'. Importantly, **a-one* 'all one' is not available in these dialects, and we do not know why this should be. We thank Jen Smith and Dominic Watt for helpful discussions of these facts. An apparently innovative *wh–* quantifier occurs in American English, where *whobody* 'who' occurs in a 1977 children's book title *Whobody there?* (Charles and Ann Morse, Upper Room Books), as well as in contemporary blogs. We are not aware of a *–one* counterpart, **who-one*. *Dogsbody*, which occurs in the ONZE data but is British in origin, is an unrelated nominal compound.

12. The Scottish Corpus of Texts and Speech is available at http://www.scottishcorpus.ac.uk/.

13. The Scottish Corpus provides only a speaker's decade of birth, not a precise year. As such, the dates listed in Table 1 have a slightly wider range than those for speakers from the other regions.

14. We used the lmer() function in the lme4 package in R (R Core Development Team, 2012).

15. We note that SES is operationalized somewhat differently across these datasets. However, our binary determination of "professional"/"nonprofessional" largely corresponds to the more traditional grouping of middle/white collar and lower/blue collar and is based on a categorization that is consistent with the construction of SES in each of the individual communities considered here. In the Newcastle and Derby datasets, "professional" and "nonprofessional" correspond to different class networks in these communities, determined largely by residential area (Docherty & Foulkes, 1999; Watt & Milroy, 1999; cf. Milroy, 1980). For the York and Scots datasets, these labels correspond to speakers' occupations. For the Toronto materials, speaker education is the primary determinant; ONZE categorizations are assessed on the basis of education and occupation (Maclagan, Gordon, & Lewis, 1999).

16. The condition number (κ) for these six predictors (four social, two linguistic) is 12.16, indicating low-to-medium collinearity (Baayen, 2008:182). Variables were selected by a step-up procedure similar

to that employed in Goldvarb (Sankoff, Tagliamonte, & Smith, 2005) and Rbrul (Johnson, 2009). Fixed main predictors improving the model significantly were added level by level. The six fixed effects in Table 2 allow for 15 possible two-way interactions. We then used this same step-up procedure to evaluate those two-way combinations where plotting suggested a possible interaction. Plotting suggested no likely interactions with >2 predictors and none emerged as significant in modeling.

17. The *C* index of concordance for this model was .926, indicating an excellent goodness of fit (Baayen, 2008:204), obtained using the somers2() function in the Hmisc package of R.

18. The corpus includes 16 speakers in the younger age cohort; *n* –*body*/–*one* = 226 (Haddican, 2008–2013).

19. These materials are compositionally similar to the UK datasets, consisting of casual conversation and comprising numerous narrative sequences.

20. The C index of concordance for this model was .932, again indicating an excellent goodness of fit (Baayen, 2008:204).

21. Crossing these two factors, each with four levels, allows for 36 unique possible pairwise interactions ($[n(n + 1)]/2$, for these ($n =$) 8 levels). Of these 36, 8 were significant at $\alpha = .05$ (*p* values uncorrected for multiple comparisons): CanE&NZE, Any&Every: $p = .015$; CanE&NZE, Every&Some: $p = .001$; CanE&BrE, Every&Some: $p = .003$; CanE&BrE, Any&Some: $p = .045$; CanE&AmE, Every&Some: $p = .024$; NZE&BrE, Any&Every: $p = .018$; NZE&BrE, Any&No: $p = .022$; NZE&BrE, Any&Some: $p = .048$.

22. Separate models (not shown here) with BrE, AmE, and CanE as reference levels show no significant differences among these samples. Our results therefore yield no strong support for the claim of Biber et al. (1999:353) that –*body* is more frequent in AmE. Instead, BrE, AmE, and CanE pattern together.

REFERENCES

Aronoff, Mark. (1976). *Word formation in generative grammar*. Cambridge: MIT Press.

Baayen, R. Harald. (2008). *Analyzing linguistic data. A practical introduction to statistics using R*. Cambridge: Cambridge University Press.

Bell, Allan. (1984). Language style as audience design. *Language in Society* 13:145–204.

Biber, Douglas, Johansson, Stig, Leech, Geoffrey, Conrad, Susan, & Finegan, Edward. (1999). *Longman grammar of spoken and written English*. Harlow: Longman.

Bolinger, Dwight. (1976). The *in*-group: *One* and its compounds. In P. A. Reich (ed.), *The second LACUS forum 1975*. Columbia: Hornbeam Press. 229–237.

BNC Consortium. (2007). The British National Corpus, version 3 (BNC XML Edition). Distributed by Oxford University Computing Services. Available at: http://www.natcorp.ox.ac.uk/.

Buchstaller, Isabelle, & D'Arcy, Alexandra. (2009). Localized globalization: A multi-local, multivariate investigation of *be like*. *Journal of Sociolinguistics* 13:291–331.

Chambers, J. K. (1995). *Sociolinguistic theory: Linguistic variation and its social significance*. Oxford: Blackwell Publishers.

Cheshire, Jenny. (2005). Age and generation-specific use of language. In U. Ammon, N. Dittmar, K. J. Mattheier, & P. Trudgill (eds.) *Sociolinguistics: An introductory handbook of the science of language and society*. 2nd ed. Berlin: Mouton de Gruyter. 1552–1563.

Docherty, Gerard J., & Foulkes, Paul. (1999). Derby and Newcastle: Instrumental phonetics and variationist studies. In P. Foulkes & G. Docherty (eds.), *Urban voices*. London: Arnold. 45–71.

Eckert, Penelope. (1999). *Linguistic variation as social practice*. Oxford: Blackwell.

Embick, David. (2008). Variation and morphosyntactic theory: Competition fractionated. *Language and Linguistics Compass* 2:59–78.

Gordon, Elizabeth, Campbell, Lyle, Hay, Jennifer, Maclagan, Margaret, Sudbury, Andrea, & Trudgill, Peter. (2004). *New Zealand English. Its origins and evolution*. Cambridge: Cambridge University Press.

Gordon, Elizabeth, Maclagan, Margaret, & Hay, Jennifer. (2007). The ONZE corpus. In J. C. Beal, K. P. Corrigan, & H. L. Moisl (eds.), *Creating and digitizing language corpora*. Vol. 2. Hampshire: Palgrave Macmillan. 82–104.

Görlach, Manfred. (1991). English as a world language—The state of the art. In M. Görlach (ed.), *Englishes. Studies in varieties of English 1984–1988*. Amsterdam: Benjamins. 10–35.

Haddican, Bill. (2008–2013). *A comparative study of language change in Northern Englishes*. Economic and Social Research Council (ESRC) of Great Britain. Res 061250033. University of York.

Heine, Bernd. (2003). Grammaticalization. In B. Joseph & R. Janda (eds.), *The handbook of historical linguistics*. Malden: Blackwell. 575–601.

Hopper, Paul, & Traugott, Elizabeth. (2003). *Grammaticalization*. 2nd ed. Cambridge: Cambridge University Press.

Jespersen, Otto. (1914). *A Modern English grammar on historical principles. Part II. Syntax*. Heidelberg: Carl Winters Universitätbuchhandlung.

Johnson, Daniel E. (2009). Getting off the GoldVarb standard: Introducing Rbrul for mixed-effects variable rule analysis. *Language and Linguistics Compass* 3:359–383.

Keller, Rudi. (1989). Invisible-hand theory and language evolution. *Lingua* 77:113–127.

Kiesling, Scott, Dilley, Laura, & Raymond, William D. (2006). The variation in conversation (ViC) project: Creation of the Buckeye Corpus of Conversational Speech. Available at: http://buckeyecorpus.osu.edu/BuckeyeCorpusmanual.pdf. Accessed November 16, 2011.

Kishimoto, Hideki. (2000). Indefinite pronouns and overt n-raising. *Linguistic Inquiry* 31:557–566.

Kroch, Anthony. (1994). Morphosyntactic variation. In K. Beals (ed.), *Papers from the 30th Regional Meeting of the Chicago Linguistics Society: Parasession on variation and linguistic theory*. Chicago: Chicago Linguistic Society. 180–201.

Kroch, Anthony, Santorini, Beatrice, & Diertani, Ariel. (2004). Penn-Helsinki Parsed Corpus of Early Modern English. Available at: http://www.ling.upenn.edu/hist-corpora/PPCEME-RELEASE-2/index.html.

————. (2010). Penn Parsed Corpus of Modern British English. Available at: www.ling.upenn.edu/hist-corpora/PPCMBE-RELEASE-1/index.html.

Kytö, Merja. (comp.) (1996). *Manual to the diachronic part of the Helsinki Corpus of English Texts: Coding conventions and lists of source text*. 3rd ed. Helsinki: Department of English, University of Helsinki. Available at: http://khnt.hit.uib.no/icame/manuals/hc/INDEX.HTM.

Labov, William. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.

————. (1990). The intersection of sex and social class in the course of linguistic change. *Language Variation and Change* 2:205–254.

————. (2001). *Principles of linguistic change: Social factors*. Oxford: Blackwell.

Larson, Richard K., & Marušič, Franc. (2004). Indefinite pronoun structures with APs: A reply to Kishimoto. *Linguistic Inquiry* 35:268–287.

Lehmann, Christian. (1982). *Thoughts on grammaticalization. A programmatic sketch*. Arbeiten der Kölner Universalien-Projeckts. Cologne: Univesität zu Köln, Institut für Sprachwissenschaft.

Maclagan, Margaret, Gordon, Elizabeth, & Lewis, Gillian. (1999). Women and sound change: Conservative and innovative behaviour by the same speakers. *Language Variation and Change* 11:19–41.

Milić, Louis T. (1990a). The Century of Prose Corpus. *Literary and Linguistic Computing* 5:203–208.

————. (1990b). A new historical corpus. *ICAME Journal* 14:26–39.

Milroy, Lesley. (1980). *Language and social networks*. Oxford: Blackwell.

Milroy, Lesley, Milroy, James, & Docherty, Gerald. (1997). *Phonological variation and change in contemporary spoken British English*. Final Report to the ESRC, R000234892. Newcastle upon Tyne: Newcastle University, Department of English Literature, Languages and Linguistics.

Milroy, Leslie, & Gordon, Matthew. (2003). *Sociolinguistics: Method and interpretation*. Malden: Blackwell.

Nevalainen, Terttu, & Raumolin-Brunberg, Helena. (2003). *Historical sociolinguistics: Language change in Tudor and Stuart England*. Harlow: Pearson Education Limited.

Nevalainen, Terttu, Raumolin-Brunberg, Helena, Keränen, Jukka, Nevala, Minna, Nurmi, Arja, & Palander-Collin, Minna. (1998). Corpus of Early English Correspondence. Helsinki: University of Helsinki, Department of English.

Pitt, Mark, Dilley, Laura, Johnson, Keith, Kiesling, Scott, Raymond, William D., Hume, Elizabeth, & Fosler-Lussier, E. (2007). Buckeye Corpus of Conversational Speech. 2nd release. Columbus: Ohio State University, Department of Psychology. Available at: www.buckeyecorpus.osu.edu.

Quirk, Randolph, Greenbaum, Sidney, Leech, Geoffrey, & Svartvik, Jan. (1985). *A comprehensive grammar of the English language*. London: Longman.

R Core Development Team. (2012). R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, Available at: http://www.R-project.org/.

Raumolin-Brunberg, Helena. (1991). *The noun phrase in early sixteenth-century English: A study based on Sir Thomas More's writings*. Mémoires de la Société Néophilologique de Helsinki 50. Helsinki: Société Néophilologique.

————. (1994). The development of the compound pronouns in –*one* and –*body* in Early Modern English. In D. Kastovsky (ed.), *Studies in Early Modern English*. Berlin: Mouton de Gruyter. 301–324.

Raumolin-Brunberg, Helena, & Kahlas-Tarkka, Leena. (1997). Indefinite pronouns with singular human reference. In M. Rissanen, M. Kytö, & K. Heikkonen (eds.), *Grammaticalization at work: Studies of long-term developments in English*. Berlin: Mouton de Gruyter. 17–85.

Rissanen, Matti. (1967). *The uses of 'one' in Old and Early Middle English*. Helsinki: Modern Language Society.

———. (1997). Whatever happened to the Middle English indefinite pronouns? In J. Fisiak (ed.), *Studies in Middle English linguistics*. Trends in Linguistics, Studies and Monographs 103. Berlin: Mouton de Gruyter. 513–529.

Rissanen, Matti, Kytö, Merja, & Palander-Collin, Minna. (eds.) (1996). *Early English in the computing age: Explorations through the Helsinki Corpus*. Berlin: Mouton de Gruyter.

Sankoff, David, Tagliamonte, Sali A., & Smith, Eric. (2005). Goldvarb X: A variable rule application for Macintosh and Windows. Toronto: University of Toronto, Department of Linguistics. Available at: http://individual.utoronto.ca/tagliamonte/Goldvarb/GV_index.htm.

Tagliamonte, Sali A. (1996–1998). *Roots of identity: Variation and grammaticalization in contemporary British English*. Economic and Social Research Council (ESRC) of Great Britain. Ref R000221842. University of York.

———. (2003–2006). *Linguistic changes in Canada entering the 21st century*. Social Sciences and Humanities Research Council of Canada (SSHRC) research grant 410-2003-0005. University of Toronto. Available at: http://individual.utoronto.ca/tagliamonte/.

———. (2006). Historical change in synchronic perspective: The legacy of British dialects. In A. van Kemenade & B. Lou (eds.), *Handbook on the history of English*. Malden: Blackwell. 477–506.

Tagliamonte, Sali A., & D'Arcy, Alexandra. (2007). Frequency and variation in the community grammar: Tracking a new change through the generations. *Language Variation and Change* 19:199–217.

Tagliamonte, Sali A., & Hudson, Rachel. (1999). Be *like* et al. beyond America: The quotative system in British and Canadian youth. *Journal of Sociolinguistics* 3:147–172.

Tagliamonte, Sali A., & Roeder, Rebecca Virginia (2009). Variation in the English definite article: Socio-historical linguistics in t'speech community. *Journal of Sociolinguistics* 13:435–471.

Taylor, Ann. (1994). Variation in past tense formation in the history of English. In R. Izvorski, M. Meyerhoff, B. Reynolds, & V. Tredinnick (eds.), *University of Pennsylvania Working Papers in Linguistics* 1:143–159.

Taylor, Ann, Nurmi, Arja, Warner, Anthony, Pintzuk, Susan, & Nevalainen, Terttu. (annotators) (2006). Parsed Corpus of Early English Correspondence, parsed version. Compiled by the CEEC Project Team. York: University of York, and Helsinki: University of Helsinki. Distributed through the Oxford Text Archive.

Venezky, Richard L., & Healey, Antonette diPaolo. (1980). *A microfiche concordance to Old English*. Publications of the Dictionary of Old English 1. Toronto: The Pontifical Institute of Mediaeval Studies.

Watt, Dominic, & Milroy, Lesley. (1999). Patterns of variation and change in three Newcastle vowels: Is this dialect leveling? In P. Foulkes & G. Docherty (eds.), *Urban voices*. London: Arnold. 25–46.

Wells, Stanley, & Taylor, Gary. (eds.) (1986). *William Shakespeare. The complete works*. Electronic ed. Oxford: Oxford University Press.

Zanuttini, R. (2008). Encoding the addressee in the syntax: Evidence from English imperatives. *Natural Language and Linguistic Theory* 28:185–218.